基于微博大数据分析时间取向与主观幸福感的关系

赵光义[1,2] 孟广腾[1,2] 井海洋[1,2] 毛浙飞[1,2] 朱廷劭[1,*]
[1]（中国科学院心理研究所 北京 100101）
[2]（中国科学院大学 北京 100049）

摘要：
人们在思考或行动上所偏好的时间方向在一定程度上影响行为，而行为会影响生活状态，据此我们希望探讨时间取向与人的主观幸福感是否有关。本研究利用微博大数据，收集了 2010 至 2017 年共 64160 名活跃用户的微博，通过关键词提取以及数据分析发现，主观幸福感与未来词频之间存在中等强度的相关（$r = 0.404, p < 0.01$），与现在词频及过去词频之间呈弱相关；将用户按时间取向词频分组后，高词频组的主观幸福感显著高于低词频组（$t = 67.442$，$p < 0.001$），同时我们也发现了主观幸福感与时间取向词频均有逐年下降的趋势。研究结果说明未来时间取向可作为主观幸福感的预测指标，本研究为利用微博大数据预测心理健康提供了新的方向。
关键词：时间取向 主观幸福感 微博 大数据 时间趋势

1 引言

个体在思考或行动时所偏好的时间方向被称为“时间取向”或“时间定向”。时间取向的形成受到社交、教育、文化等因素的影响，与诸多人格特质的形成有着密切的关系，在个体认知和理解现实生活经验以及形成动机决策方面均发挥重要作用。主观幸福感主要是指人们对其生活质量所做的情感性和认知性的整体评价，可用来评估相当长一段时期的情感反应和生活满意度，人格因素是其重要的预测指标。本研究利用微博大数据，通过文本分析获取用户的时间取向词频与主观幸福感，通过相关性分析，探究时间取向与主观幸福感的关系。

1.1 时间取向

人们常对某一特定时间延伸方向存在思维上的认知定向或偏向，如有人习惯于回顾过往、沉湎于过去，有人看重现在、活在当下，也有人强调将来、憧憬未来，这种个体在思考或行动时所偏好的时间方向称为“时间取向”或“时间定向”[1]。在时间取向的形成中，不仅社交、模仿、教育、文化以及其他环境因素起了重要作用[2]，一些动态的情境性因素，如娱乐、自满、压力等也影响时间取向的塑成[3, 4]。当时间取向形成后，其对个体的情感，认知等方面均起着重大作用。

近些年来，研究发现时间取向与某些心理特质有着密切的关系。例如，在与拖延的相关研究中，Specter 和 Ferrari [5]表明拖延与未来时间取向成负相关，而与过去时间取向成正相关，与现在时间取向无显著相关。还有研究表明，时间取向还在自我效能[6]、自我调节[7]、延迟满足[8]等人格特质中发挥重要作用。国内的研究也从认知建构层面上发现时取向对个体自信建构存在影响，未来时间取向组的自信判断显著高于过去和现在时间取向组[1]。

* 通讯作者：朱廷劭，Email：tszhu@psych.ac.cn

1.2 主观幸福感

主观幸福感(Subjective Well-Being, SWB)是专指评价者根据自定的标准对其生活质量的整体性评估，它是衡量个人生活质量的重要的综合性心理指标。Diener[9] 认为主观幸福感具有主观性、相对稳定性和整体性三个特点。国内外有很多关于主观幸福感的研究，也制定了许多测量主观幸福感的量表，如国外广泛使用的 Argyle 编制的 29 题牛津幸福感问卷（修订版），该问卷采用 0～3 的四级计分，只有 1 个维度，将各题得分相加即为幸福感总分，总分越高代表个体的幸福感越[10]。

1.3 微博与心理健康

微博是一个由新浪网推出并提供微型博客服务类的社交网站。用户可以通过网页、WAP 页面、手机客户端、手机短信、彩信发布消息或上传图片。用户可以将看到的、听到的、想到的事情写成一句话，或发一张图片，通过电脑或者手机随时随地分享给朋友，一起分享、讨论；还可以关注朋友，即时看到朋友们发布的信息。近年的研究表明，使用微博这类社交媒体的强度与个体的生活满意度存在相关，社交媒体的使用对个体的积极情绪与消极情绪都有影响[11]。

2 方法

2.1 变量

我们探究时间取向和主观幸福感之间的关系，以时间取向为自变量，主观幸福感为因变量，同时也考虑了年份等变量是否与因变量相关。主要变量的数据采集方法如下：

时间取向 我们统计了每位微博用户每年微博中的时间取向词频，通过爬取微博文本，使用中科院心理研究所计算网络实验室开发的“文心（TextMind）”进行分析，我们能获取指定文本中的关键词词频。文心所使用的中文词库参照了经典的 LIWC2007 和正體中文 C-LIWC 词库[12]。LIWC 因其较好的信效度，问世 20 多年来被心理学研究者广泛使用，C-LIWC 则是其中文繁体版，最早由台湾学者黄金兰等开发和完善[13]。而文心则是基于简体中文背景下的中文文本分析[14]，最适合本文的研究。

本研究中，时间取向词分为过去、现在、未来三种：过去词包含“曾经”、“从前”、“昨天”、“上周”、“当初”等 24 个词；现在词包含“此刻”、“本周”、“今夜”、“目前”等 20 个词；未来词包括“以后“、”即将“、”来世“、”改天”等 20 个词，详见附表 1。我们将微博文本分析放入文心系统，文心逐行对文本进行分词，并将不同类型的词（如“上周”、“从前”）归纳为同一特征（如过去），进而得到微博文本中时间取向词数。凡是微博中含有上述词汇的，均累加计数。为使得时间取向词更有代表性，我们使用词频作为实际分析中的自变量，即过去词频=(过去词数)/微博数量，现在词频=(现在词数)/微博数量，未来词频=(未来词)/微博数量；同时引入“时间取向词频”也作为自变量变量，时间取向词频=(过去词+现在词+未来词)/微博数量。

主观幸福感 主观幸福感来源于中国科学院心理研究所计算网络实验室的研究模型。其基本原理是在个体层面上对社会媒体用户进行主观幸福感的预测，通过划分不同的特征集合：人口统计学信息、用户微博行为特征、用户微博内容特征，以其主观幸福感作为目标变量，通过 StepWise、LASSO、Support Vector Regression、Multivariate Adaptive Regression Splines 等机器学习算法来进行建模，取得的预测效果能够达到其他研究者以居民人均收入、受教育水平等客观统计指标达到的最佳预测准确度水平，预测模型得到的预测分数与心理学量表得到的测评分数之间的相关系数达到 0.6。本研究中，我们使用该实验室的成型模型计算每名微博用户的主观幸福感得分。

2.2 被试

我们使用的数据来源于 10000 名微博用户在 2010 年 1 月 1 日至 2017 年 12 月 31 日之间所发的、文本描述中涉及到关键词的有效微博，总计 8 年。因为我们的数据按照年份进行爬取，统计每年均值，而不同年份之间会有同一名用户的数据，我们也记为新用户，以此探索变量的时间变化趋势。数据清理阶段，删去了微博数量，主观幸福感，时间取向总词频在三个标准差之外的数据，最终共有 64160 条数据被纳入分析。

2.3 统计方法

主要采用 SPSS 24.0 进行数据处理。首先对变量进行描述性分析，计算均值与标准差。接着对主观幸福感与各时间取向词频之间进行相关分析。根据初步分析的结果，我们将被试按时间取向词频中位数分组，对两组主观幸福感得分进行独立样本 t 检验。最后，我们考虑时间趋势的影响，分不同年份进行了 t 检验和相关分析。

3 结果

对选定的变量进行简单的描述统计，删除微博数量、时间取向词频、主观幸福感数值在平均值加减三个标准差之外的被试，获得最终用于分析的被试共 64160 人，对被试各变量的描述性统计结果见表 1。2010 年至 2017 年，微博用户的时间取向词频为 0.104±0.094，即平均约十篇微博中出现一次时间取向词，其中现在词出现的频率较高，为 0.052±0.059，过去词频次之，为 0.031±0.039，未来词频最低，为 0.020±0.029。

表 1 微博用户主观幸福感与时间取向词频统计（N＝64160）

	平均值	标准差
主观幸福感	103.294	29.293
时间取向词频	.104	.094
过去词频	.031	.039
现在词频	.052	.059
未来词频	.020	.029

为探究主观幸福感与三种时间取向词频之间的关系，接下来对这些变量两两间进行相关分析，由于这些变量均不服从正态分布，表 2 中报告的均为斯皮尔曼

相关系数。相关分析显示，主观幸福感与未来词频之间达到中等强度相关（r = 0.404, p < 0.01），与现在词频与过去词频之间呈弱相关，与时间取向词频也为弱相关。此外，时间取向词频与三种时间取向词频均有显著的强相关，而三种时间取向词频之间也呈现中等强度的相关。

表 2 主观幸福感与时间取向词频的相关性

	时间取向词频	未来词频	现在词频	过去词频
未来词频	.697**			
现在词频	.883**	.549**		
过去词频	.766**	.552**	.560**	
主观幸福感	.326**	.404**	.329**	.374**

** p < .001

由于现有数据下将被试按三种时间取向分组较为困难，而各时间取向词的出现频率均与时间取向词频高相关，且频数分析发现有 16.3%的微博用户的时间取向词频为 0，即微博中无时间取向词出现，因此我们希望进一步研究时间取向词频与主观幸福感的关系，于是以时间取向词频中位数为截点，将用户分为高时间取向词频组（简称高词频组）与低时间取向词频组（简称低词频组），进行独立样本 t 检验，结果如图 1 所示，高词频组的主观幸福感显著高于低词频组（110.81±32.89 v.s. 95.75±22.81, t = 67.442 , p < 0.001）。又考虑到时间的影响，遂将时间取向词频中位数逐年计算并按同样的方式分组，逐年进行独立样本 t 检验，结果如图 2 所示，高词频组的主观幸福感始终显著高于低词频组（p < 0.01）。

对微博用户每年的过去词频、现在词频、未来词频进行 ANOVA 分析，结果表明每年的不同时间取向词频间均存在显著差异（Fs > 475.932, ps < .001），事后分析表明，现在词频显著高于过去词频，过去词频显著高于未来词频（ps < .001）。同样，主观幸福感逐年间也存在显著差异（ps < 0.001）。对主观相关感和各时间取向词频以及总词频进行相关分析，具体相关系数见表 3。

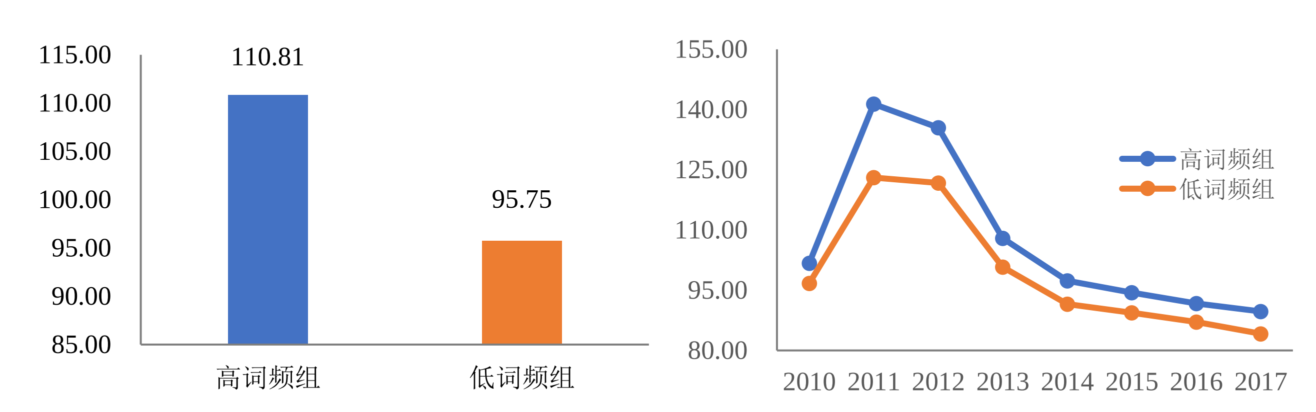


图 2 高低时间词频组主观幸福感比较　　**图 1 2010-2017 年高低时间词频组主观幸福感比较**

表 3 主观幸福感与各时间取向词频的描述与相关统计 [a]

年份	N	主观幸福感	时间取向词频	过去词频	现在词频	未来词频
2010	4750	99.222 ± 19.947	.138 ± .108	.038 ± .045	.073 ± .070	.027 ± .033
			.228**	.298**	.275**	.270**
2011	7658	132.214 ± 37.872	.118 ± .080	.035 ± .028	.059 ± .046	.025 ± .024
			.295**	.317**	.317**	.275**
2012	8557	128.612 ± 35.297	.125 ± .077	.037 ± .027	.062 ± .044	.026 ± .020
			.248**	.235**	.251**	.265**
2013	9057	104.361 ± 23.034	.121 ± .084	.036 ± .033	.060 ± .050	.025 ± .025
			.218**	.251**	.203**	.277**
2014	8998	94.461 ± 18.655	.102 ± .096	.032 ± .042	.051 ± .063	.019 ± .030
			.188**	.248**	.172**	.269**
2015	8529	91.877 ± 16.614	.091 ± .098	.028 ± .045	.047 ± .067	.017 ± .034
			.180**	.261**	.167**	.251**
2016	8151	89.398 ± 16.721	.083 ± .096	.026 ± .044	.042 ± .062	.016 ± .033
			.165**	.216**	.146**	.205**
2017	8460	86.945 ± 12.517	.070 ± .097	.022 ± .044	.035 ± .062	.013 ± .030
			.287**	.279**	.255**	.241**

a. 每年数据的第一行为该变量的均值±标准差，第二行为该变量与主观幸福感的斯皮尔曼相关系数

** $p < .001$

4 讨论

时间取向词与情绪、人格等心理特征有一定的联系，我们的研究显示，2010 年至 2017 年时间取向词在微博中出现的频率平均为 0.104 次，其中现在词出现最多，过去词、未来词次之。微博作为表达个人状态与讨论实事热点的重要平台，与“现在”相关的词出现次数是最多的，我们的结果与常识相符。而总体的相关分析则显示，未来词频与主观幸福感的相关性最高，达到中等相关强度，未来词的使用可以在一定程度上反映用户主观幸福感的情况。

按照时间取向词频进行高低分组后的分析显示，使用时间取向词的频率也会影响主观幸福感水平，高词频组要显著高于低词频组。这可能因为主观幸福感高的个体更愿意表达时间相关的内容，这也在一定程度上解释了现在词频与过去词频也均与主观幸福感存在正相关的现象。

从表 3 中可以发现，微博用户的主观幸福感在 2010-2011 年骤升，之后处于持续下降的水平。但由于这里使用的主观幸福感并非来自直接测量，而是通过机器学习计算得出的，所以还需要进一步的研究证据。此外，可以看出过去词频、现在词频、未来词频三者的也在 2012-2017 年持续下降，这说明微博用户在近年来提及时间取向词的频率越来越低。这一现象可能的解释是，2010 年至 2012 年，微博的用法更接近于博客，发布和浏览微博严重依赖电脑。但 2013 年后，随着智能手机和微博客户端的普及，用户可以随时随地发布和浏览微博，提高了微博

的即时性，从而不再需要通过文字提示微博记录事件的发生时间。并且，随着朋友圈等其他平台的兴起，微博作为日常事件记录的功能逐渐降低，用户更倾向于发布一些当下没有明确时间点的心情和状态。

通过相关分析发现，现在词频与主观幸福感的相关系数总是略低于过去词频和未来词频，而过去词频和未来词频之间在多年的数据中基本持平，总体处于较低水平的正相关。这表明无论是提及过去、现在还是未来，微博用户更倾向于呈现出积极的状态，表现出更高水平的主观幸福感。

5 结论与展望

本研究利用微博大数据，探讨了时间取向与主观幸福感之间的关系，并分析了主观幸福感与时间取向词频随年份的变化趋势，为利用微博大数据预测心理健康提供了新的方向。由于微博数据的限制，我们的分析未纳入年龄、性别、地区、社会经济低位等人口学因素，同时，主观幸福感得分是通过模型计算得出的，其准确性不及直接测量，因此本研究的结果有待进一步验证。以后的研究可以采用问卷调查等方式，收集更全面准确的数据，分析时间取向与主观幸福感的关系。

参考文献

[1] 狄轩康. 时间取向和时间态度对自信建构的影响 [D]; 西南大学, 2014.

[2] Seginer R. UNIT 11, CHAPTER 5 ADOLESCENT FUTURE ORIENTATION: AN INTEGRATED CULTURAL AND ECOLOGICAL PERSPECTIVE [J]. Wwu Edu, 1970.

[3] Beiser M. Changing time perspective and mental health among Southeast Asian refugees [J]. Culture Medicine & Psychiatry, 1987, 11(4): 437-64.。

[4] Zimbardo P G, Boyd J N. Putting Time in Perspective: A Valid, Reliable Individual-Differences Metric [J]. Journal of Personality & Social Psychology, 1999, 77(6): 1271-88.

[5] Specter M H, Ferrari J R. Time Orientations of Procrastinators--Focus on the Past, Present, or Future? [J]. Journal of Social Behavior & Personality, 2000, 15(5): 197-202.

[6] Zebardast A, Besharat M A, Hghighatgoo M. The Relationship between Self-Efficacy and Time Perspective in Students [J]. Procedia - Social and Behavioral Sciences, 2011, 30(912-5.

[7] Zebardast A, Besharat M A, Hghighatgoo M. The Relationship between Self-Regulation and Time Perspective in Students [J]. 2011, 30(935-8.

[8] Daugherty J R, Brase G L. Taking time to be healthy: Predicting health behaviors with delay discounting and time perspective [J]. Personality & Individual Differences, 2010, 48(2): 202-7.

[9] Diener E. Subjective Well-Being [J]. Psychological Bulletin, 1984, 95(3): 542-75.

[10] 李义安, 陈彦垒. 牛津幸福感问卷(修订版)的维度结构及信效度检验 [J]. 保健医学研究与实践, 2013, 10(01): 34-7+41.

[11] 唐嘉仪. 社交媒体使用与大学生幸福感关系：对比 QQ 空间和新浪微博 [D]; 中山大学, 2014.

[12] 吴育锋, 吴胜涛, 朱廷劭, et al. 小说人物性格的文学智能分析:以《平凡的世界》为例 [J]. 中文信息学报, 2018, 32(07): 128-36.

[13] 张信勇. LIWC:一种基于语词计量的文本分析工具 [J]. 西南民族大学学报(人文社会科学版), 2015, 36(04): 101-4.

[14] Rui G, Hao B, He L, et al. Developing Simplified Chinese Psychological Linguistic Analysis Dictionary for Microblog [M]. 2013.

作者贡献声明：

赵光义：提出研究思路，数据清洗及分析，论文摘要、结果、讨论部分撰写
孟广腾：设计研究方案，数据清洗及分析，论文结果、讨论部分撰写
井海洋：文献收集，论文背景部分撰写
毛浙飞：文献收集，论文方法部分撰写
朱廷劭：研究思路指导以及过程管理

附表 1 时间取向关键词

时间取向	关键词
过去	'曾经', '从前', '当初', '当年', '当时', '古代', '过去', '过往', '前年', '前世', '前天', '去年', '往日', '昔日', '先前', '已经', '以前', '以往', '原先', '昨日', '昨天', '上周', '上星期', '上礼拜'
现在	'此刻','此时', '当前', '当天', '当下', '即刻', '今日', '今天', '今晚', '今夜', '近期', '目前', '如今', '现阶段', '现今', '现时', '现在', '正在', '本周', '本星期'
未来	'大后天', '改天', '后天', '即将', '将来', '将要', '接下来', '今后', '来生', '来世', '明年', '明日', '明天', '日后', '往后', '未来', '以后', '下周', '下星期', '下礼拜'